



HAL
open science

On the genericity of Whitehead minimality

Frédérique Bassino, Cyril Nicaud, Pascal Weil

► **To cite this version:**

Frédérique Bassino, Cyril Nicaud, Pascal Weil. On the genericity of Whitehead minimality. 2013.
hal-00919489v1

HAL Id: hal-00919489

<https://hal.science/hal-00919489v1>

Submitted on 16 Dec 2013 (v1), last revised 6 Mar 2014 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the genericity of Whitehead minimality^{*}

Frédérique Bassino¹, Cyril Nicaud², and Pascal Weil^{3,4}

¹ Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS, (UMR 7030), F-93430, Villetaneuse, France. bassino@lipn.univ-paris13.fr

² Université Paris-Est, LIGM, CNRS UMR 8049, F-77454 Marne-la-Vallée, France. nicaud@univ-mlv.fr

³ CNRS, LaBRI, UMR 5800, F-33400 Talence, France, pascal.weil@labri.fr

⁴ Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France.

Abstract. We show that a finitely generated subgroups of a free group, chosen uniformly at random, is strictly Whitehead minimal. Whitehead minimality is one of the key elements of the solution of the orbit problem in free groups. The proofs strongly rely on automata theory and on the combinatorial tools used in the analysis of algorithms. The result we prove actually depends implicitly on the choice of a distribution on finitely generated subgroups, and we establish it for the two distributions which appear in the literature on random subgroups.

1 Introduction

The contribution of this paper is the solution of a problem in group theory, by methods arising from theoretical computer science – specifically: combinatorial and probabilistic methods that were developed to analyze the performance of algorithms [5].

The problem in question is the generic complexity of the *Whitehead minimization problem* for finitely generated subgroups of a free group $F(A)$. Every such subgroup H is a rational subset of $F(A)$ and can be represented uniquely by a finite state automaton $\Gamma(H)$ with particular constraints, called the Stallings graph of the subgroup; this automaton naturally constitutes a privileged tool to compute with subgroups, and it also provides a notion of size for H : we denote by $|H|$ the number of states of $\Gamma(H)$.

A natural equivalence relation on subgroups is provided by the action of the automorphism group of $F(A)$: the subgroups H and K are in the same orbit if $K = \varphi(H)$ for some automorphism φ of $F(A)$ — that is, H and K are “the same” up to a change of basis in the ambient group. The Whitehead minimization problem consists in deciding whether H is a minimum size representative of its orbit (we say that H is Whitehead minimal). We refer the readers to [15] for a (worst-case) polynomial time algorithm deciding the Whitehead minimization problem, based on an early result of Gersten [6] and to [12] for the usage of this problem in solving the more general orbit problem.

^{*} This work was partially supported by the ANR through ANR-2010-BLAN-0204, through ANR-10-LABX-58 and through ANR-JCJC-12-JS02-012-01

Here we are rather interested in the notion of generic complexity, that is, the complexity of the problem when restricted to a generic set of instances (a set of instances such that an instance of size n sits in it with probability tending to 1 when n tends to infinity; precise definitions are given below). Our main result states that the generic complexity of the Whitehead minimization problem is constant, and more precisely, that the set of Whitehead minimal subgroups is generic (see [13] for an early discussion of generic complexity, specially in the case of cyclic subgroups).

An implicit dimension of the discussion of complexity is the notion of size of inputs. In the case of finitely generated subgroups of a free group, we can use either a k -tuple (k fixed) of words which are generators of the subgroup H (and the size of the input is the sum of the lengths of these words), or the Stallings graph of H (and the size is $|H|$). These two ways of specifying the subgroup H give closely related worst-case complexities (because of linear inequalities between the two notions of size), but they can give very different generic complexities: it was shown in [2] that malnormality is generic if subgroups are specified by a tuple of generators, whereas non-malnormality is generic if subgroups are specified by their Stallings graph. Our results show that Whitehead minimality is generic in both set-ups.

A key ingredient of our proof is a purely combinatorial characterization of Whitehead minimality in terms of the properties of the automaton $\Gamma(H)$ (Proposition 1 below), proved in [15], which involves counting the transitions labeled by certain subsets of the alphabet in and out of each state. This is what allows us to turn the algebraic problem into a combinatorial one, which can be tackled with the methods of combinatorics and theoretical computer science.

Interestingly, the reasons why Whitehead minimality is generic when subgroups are specified by their Stallings graph, and why it is generic when subgroups are specified by a k -tuple of words, are diametrically opposed. The Stallings graph of the subgroup generated by a k -tuple of words of length at most n generically consists of a small central tree and long loops connecting leaves of the tree, so much of the geometry of the graph is along these long loops, where each state is adjacent to only two transitions. In contrast, an n -vertex Stallings graph generically has many transitions and each state is adjacent to a near-full set of transitions.

The origins of this work go back to discussions with Armando Martino and Enric Ventura in 2009.

2 Preliminaries

Let $r > 1$, let A be a finite r -element set and let $F(A)$ be the *free group on A* . We can think of $F(A)$ as the set of *reduced* words on the symmetrized alphabet $\tilde{A} = A \cup \bar{A}$, where $\bar{A} = \{\bar{a} \mid a \in A\}$. Recall that a word is reduced if it does not contain occurrences of the words of the form $a\bar{a}$ or $\bar{a}a$ ($a \in A$). The operation $x \mapsto \bar{x}$ is extended to \tilde{A} by letting $\bar{\bar{a}} = a$.

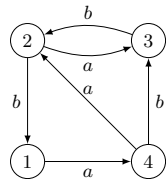


Fig. 1. The Stallings graph of $H = \langle aab, ab\bar{a}b, abbb \rangle$. The reduced word $u = aab\bar{a}b$ is in H as it is accepted by $\Gamma(H)$: it labels a path starting from 1 and ending at 1, with edges being used backward when reading a negative letter. Since every vertex has valency at least 2, this graph is cyclically reduced.

We denote by $[n]$ the set of positive integers less than or equal to n , and by \mathcal{R}_n the set of reduced words of length at most n . A reduced word u is called *cyclically reduced* if u^2 is reduced, and we let \mathcal{C}_n be the set of cyclically reduced words of length at most n .

2.1 Stallings graph

It is classical to represent the finitely generated subgroups of a free group by a finite state automaton, subject to certain combinatorial constraints. An *A-graph* is a finite graph Γ whose edges are labeled by elements of A . It can be seen also as a transition system on alphabet \bar{A} , with the convention that every a -edge from p to q represents an a -transition from p to q and an \bar{a} -transition from q to p . Say that Γ is *reduced* if it is connected and if no two edges with the same label start (resp. end) at the same vertex: this is equivalent to stating that the corresponding transition system is deterministic and co-deterministic. If 1 is a vertex of Γ , we say that $(\Gamma, 1)$ is *rooted* if every vertex, except possibly 1, has valency at least 2.

We say that Γ is *cyclically reduced* if it is reduced and every vertex has valency at least 2. The A -graph in Fig. 1 is cyclically reduced.

If H is a finitely generated subgroup of $F(A)$, there exists a unique reduced rooted graph $(\Gamma(H), 1)$, called the *Stallings graph* of H , such that H is exactly the set of reduced words accepted by $(\Gamma(H), 1)$: a reduced word is accepted when it labels a loop starting and ending at 1. Moreover, this graph can be effectively computed given a tuple of reduced words generating H , in time $\mathcal{O}(n \log^* n)$ [16, 17]. We denote by $|H|$ the number of vertices of $\Gamma(H)$, which we interpret as a notion of “size” of H .

2.2 Whitehead minimality

As explained in the introduction, a subgroup H is *Whitehead minimal* if it has minimum size in its automorphic orbit, that is if $|H| \leq |\varphi(H)|$ for every automorphism φ of $F(A)$. It is *strictly Whitehead minimal* if $|H| < |\varphi(H)|$ for every automorphism φ that is not length preserving (i.e., that is not induced by a permutation of \bar{A}). Strict Whitehead minimality means that H is the only minimum size representative of its orbit, up to a permutation of the letters (that is, up to a relabeling of the edges of its Stallings graph).

A crucial characterization of (strict) Whitehead minimality can be expressed in terms of the so-called Whitehead automorphisms [6]. In this paper we rather

present directly a derived characterization, in terms of the combinatorial properties of $\Gamma(H)$, proved in [15].

A pair (Y, v) consisting of a subset Y of \tilde{A} and a letter $v \in \tilde{A}$, is called a *Whitehead descriptor* if $v \in Y$, $\bar{v} \notin Y$ and $2 \leq |Y| \leq 2|A| - 2$. There is a bijection between the non-length preserving Whitehead automorphisms and these descriptors, see for instance [15].

Let Γ be a reduced graph, and let (Y, v) be a Whitehead descriptor. Then we let $\mathbf{positive}(\Gamma, Y, v)$ be the set of vertices of Γ with at least one incoming edge labeled by a letter in Y , at least one incoming edge labeled by a letter not in Y , and no incoming edge labeled v . Let also $\mathbf{negative}(\Gamma, Y, v)$ be the set of vertices with an incoming edge labeled v , and all other incoming edges labeled by letters in Y .

Example. Consider the Whitehead descriptor (Y, v) with $v = \bar{a}$ and $Y = \{\bar{a}, b\}$. For the graph Γ depicted on Fig. 1, vertex 1 is in $\mathbf{negative}(\Gamma, Y, v)$ since its incoming edges are labeled by b and \bar{a} (obtained by flipping the edge $1 \xrightarrow{a} 4$). Vertex 3 is in $\mathbf{positive}(\Gamma, Y, v)$ since its incoming edges are labeled by a and b , one not in Y , one in Y and both different from v . One can also verify that vertices 2 and 4 are neither in $\mathbf{positive}(\Gamma, Y, v)$ nor in $\mathbf{negative}(\Gamma, Y, v)$.

The following statement is an immediate consequence of [15, Proposition 2.4 and Remark 2.6].

Proposition 1. *A finitely generated subgroup H of $F(A)$ is strictly Whitehead minimal if and only if it is cyclically reduced and, for every Whitehead descriptor (Y, v) , we have $|\mathbf{positive}(\Gamma(H), Y, v)| > |\mathbf{negative}(\Gamma(H), Y, v)|$.*

2.3 Distributions over finitely generated subgroups

Complexity. Let S be a countable set, the disjoint union of finite sets S_n ($n \geq 0$), and let $B_n = \bigcup_{i \leq n} S_i$. Typically in this paper, S will be the set of Stallings graphs, of partial injections, of reduced words or of k -tuples of reduced words, and S_n will be the set of elements of S of size n .

A subset X of S is *negligible* if the probability for an element of B_n to be in X , tends to 0 when n tends to infinity; that is, if $\lim_n \frac{|X \cap B_n|}{|B_n|} = 0$.

The notion is refined as follows: we say that X is *exponentially* (resp. *super-polynomially*, *polynomially*) *negligible* if $\frac{|X \cap B_n|}{|B_n|}$ is $\mathcal{O}(e^{-cn})$ for some $c > 0$ (resp. $\mathcal{O}(n^{-k})$ for every positive integer k , $\mathcal{O}(n^{-k})$ for some positive integer k). The set X is *exponentially* (resp. *super-polynomially*, *polynomially*, *simply*) *generic* if its complement is *exponentially* (resp. *super-polynomially*, *polynomially*, *simply*) *negligible*. We note the following elementary lemma.

Lemma 1. *With the above notation, if $C \subseteq S$ satisfies $\liminf_n \frac{|C \cap B_n|}{|B_n|} = p > 0$ and X is *exponentially* (resp. *super-polynomially*, *polynomially*, *simply*) *negligible* in S , then so is $X \cap C$ in C .*

Proof. The verification is immediate if we observe that, for n large enough,

$$\frac{|X \cap C \cap B_n|}{|C \cap B_n|} \leq \frac{|X \cap B_n|}{|C \cap B_n|} = \frac{|X \cap B_n|}{|B_n|} \frac{|B_n|}{|C \cap B_n|} \leq \frac{2}{p} \frac{|X \cap B_n|}{|B_n|}. \quad \square$$

Genericity and negligibility can also be defined using the radius n spheres S_n instead of the balls B_n . The same properties are generic or negligible, exponentially, super-polynomially, polynomially or simply, provided $|B_n|$ grows fast enough, see for instance [2, Sec. 2.2.2].

The graph-based distribution. The uniform distribution on the set of size n Stallings graphs was analyzed by Bassino, Nicaud and Weil [3]. Here we summarize the principles of this distribution and the features which will be used in this paper.

In a Stallings graph, each letter labels a partial injection on the vertex set: in fact, such a graph can be viewed as an A -tuple $\mathbf{f} = (f_a)_{a \in A}$ of partial injections on an n -element set, with a distinguished vertex, and such that the resulting graph (with an a -labeled edge from i to j if and only if $j = f_a(i)$) is connected and has no vertex of valency 1, except perhaps the distinguished vertex. We may even assume that the n -element set in question is $[n]$, with 1 as the distinguished vertex, see [3, Section 1.2] for a precise justification.

Let \mathcal{I}_n denote the set of partial injections on $[n]$ and let \mathcal{B}_n be the set of r -tuples in \mathcal{I}_n^r which define a Stallings graph (recall that $|A| = r$). Let also \mathcal{D}_n be the subset of \mathcal{B}_n , of those r -tuples which define a cyclically reduced Stallings graph. Then \mathcal{D}_n (and hence \mathcal{B}_n) is generic in \mathcal{I}_n^r [3, Corollary 2.7].

The fundamental observation, used in [3] to achieve this result, is the following: the functional graph of a partial injection $f \in \mathcal{I}_n$ (that is: the pair $([n], E)$ where $i \rightarrow j \in E$ whenever $j = f(i)$), is made of cycles and sequences. This allows the use of the analytic combinatorics calculus on exponential generating series (EGS) [5, Sec. II.2]. Recall that, if I_n is the number of partial injections on $[n]$, the corresponding EGS is $I(z) = \sum_{n \geq 0} \frac{I_n}{n!} z^n$. From [3, Sec. 2.1 and Proposition 2.10], we get

$$I(z) = \frac{1}{1-z} \exp\left(\frac{z}{1-z}\right) \quad \text{and} \quad \frac{I_n}{n!} = \frac{e^{-\frac{1}{2}}}{2\sqrt{\pi}} e^{2\sqrt{n}} n^{-\frac{1}{4}} (1 + o(1)). \quad (1)$$

The formula for $I(z)$ is based on the fact that a partial injection is a set of sequences (whose EGS is $\frac{z}{1-z}$) and of cycles (whose EGS is $\log\left(\frac{1}{1-z}\right)$). We refer the readers to [5, Sec. II.2] and [3] for further details. We use again this calculus in Section 3.1.

The word-based distribution. The distribution more commonly found in the literature (e.g. [10, 8, 9]), which we term *word-based*, originated in the work of Arzhantseva and Ol'shanskiĭ [1]. It is in fact a distribution on the k -tuples $\mathbf{h} = (h_1, \dots, h_k)$ of reduced words of length at most n , where k is fixed and n is allowed to grow to infinity; one then considers the subgroup H generated by \mathbf{h} .

This is a reasonable way of defining a distribution on finitely generated subgroups of $F(A)$, and even on rank k subgroups, in spite of the fact that different tuples may generate the same subgroup (see for instance [2, Sec. 3.1]).

The literature also considers the Gromov's so-called density model, which uses much larger random tuples (of positive density within \mathcal{C}_n). This model is usually considered to study the asymptotic properties of finite group presentations rather than subgroups of $F(A)$ and we will not discuss it here (see for instance [14]).

We will use the following statistics on \mathcal{R}_n and \mathcal{C}_n , which can be easily verified: $|\mathcal{R}_n| = \frac{r}{r-1}((2r-1)^n - 1)$ and $2r((2r-1)^{n-1} - 1) \leq |\mathcal{C}_n| \leq |\mathcal{R}_n|$. In particular, both $|\mathcal{R}_n|$ and $|\mathcal{C}_n|$ are $\Theta((2r-1)^n)$ and $\liminf_n \frac{|\mathcal{C}_n|}{|\mathcal{R}_n|} > 0$.

3 The graph-based distribution

We now study the genericity of strict Whitehead minimality for the graph-based distribution. The proof of Theorem 1 below is given in Sections 3.1 and 3.2.

Theorem 1. *Strict Whitehead minimality is super-polynomially generic for the uniform distribution over the set of cyclically reduced Stallings graphs.*

3.1 Statistical properties of size n partial injections

If f is a partial injection on $[n]$, we let

- $\text{sequence}(f)$ be the number of sequences in the functional graph of f ; a sequence has at least one vertex;
- $\text{extr}(f) = \{i \in [n] \mid f(i) \text{ is undefined or } i \text{ has no preimage by } f\}$; it is the set of extremities of sequences in the digraph.

We note that, for every $f \in \mathcal{I}_n$,

$$\text{sequence}(f) \leq |\text{extr}(f)| \leq 2 \text{sequence}(f). \quad (2)$$

Proposition 2. *For the uniform distribution, the probability that the number of sequences of a size n partial injection is not in $(\frac{1}{2}\sqrt{n}, 2\sqrt{n})$ is super-polynomially small (of the form $\mathcal{O}(e^{-c\sqrt{n}})$ for some $c > 0$).*

Proof. If $T(z)$ is a formal power series, we denote by $[z^n]T(z)$ the coefficient of z^n in the series. For any $k \geq 0$, let $S^k(z)$, $S^{\leq k}(z)$ and $S^{\geq k}(z)$ be the EGSs of the partial injections having respectively exactly k , at most k and at least k sequences. Observe that an injection with k sequences is a set of k sequences together with a set of cycles; the symbolic method [5, Sec. II.2] therefore yields:

$$S^k(z) = \frac{1}{k!} \left(\frac{z}{1-z} \right)^k \frac{1}{1-z}.$$

The radius of convergence of this series is 1, and Cauchy's estimate for the coefficient of a power series states that for any positive real $\zeta < 1$, we have

$$[z^n]S^k(z) \leq \frac{S^k(\zeta)}{\zeta^n}.$$

Taking $\zeta = 1 - \frac{1}{\sqrt{n}}$ approximatively minimizes the right hand quantity, and after basic computations we obtain that for n large enough,

$$[z^n]S^k(z) \leq \sqrt{n} e^{2+\sqrt{n}} \cdot \frac{n^{\frac{k+1}{2}}}{k!}.$$

Since $S^{\leq \frac{1}{2}\sqrt{n}}(z) = \sum_{k=0}^{\frac{1}{2}\sqrt{n}} S^k(z)$ and $S^{\geq 2\sqrt{n}}(z) = \sum_{k=2\sqrt{n}}^n S^k(z)$ we get upper bounds for coefficients of both series by bounding $\sum_{k=0}^{\frac{1}{2}\sqrt{n}} \frac{1}{k!} n^{\frac{k}{2}}$ and $\sum_{k=2\sqrt{n}}^n \frac{1}{k!} n^{\frac{k}{2}}$ from above. The term $\frac{1}{k!} n^{\frac{k}{2}}$ is increasing in the first sum and decreasing in the second one, so we can bound each term of each series by its maximum value. This yields the following inequalities:

$$\begin{aligned} \sum_{k=0}^{\frac{1}{2}\sqrt{n}} \frac{n^{\frac{k}{2}}}{k!} &\leq \sum_{k=0}^{\frac{1}{2}\sqrt{n}} \frac{n^{\frac{1}{2} + \frac{1}{4}\sqrt{n}}}{(\frac{1}{2}\sqrt{n})!}, & [z^n]S^{\leq \frac{1}{2}\sqrt{n}}(z) &\leq \frac{n^{\frac{3}{2} + \frac{1}{4}\sqrt{n}}}{(\frac{1}{2}\sqrt{n})!} e^{2+\sqrt{n}} \quad \text{and} \\ \sum_{k=2\sqrt{n}}^n \frac{n^{\frac{k}{2}}}{k!} &\leq \sum_{k=2\sqrt{n}}^n \frac{n^{\frac{1}{2} + \sqrt{n}}}{(2\sqrt{n})!}, & [z^n]S^{\geq 2\sqrt{n}}(z) &\leq \frac{n^{2+\sqrt{n}}}{(2\sqrt{n})!} e^{2+\sqrt{n}}. \end{aligned}$$

Using the crude Stirling bound $n! \geq n^n e^{-n}$ and the asymptotics of I_n in Eq. (1), we obtain upper bounds of the announced form for

$$\frac{[z^n]S^{\leq \frac{1}{2}\sqrt{n}}(z)}{[z^n]I(z)} \quad \text{and} \quad \frac{[z^n]S^{\geq 2\sqrt{n}}(z)}{[z^n]I(z)},$$

respectively the probabilities for a partial injection on $[n]$ to have at most $\frac{1}{2}\sqrt{n}$ and at least $2\sqrt{n}$ sequences. \square

We use Proposition 2 to bound the number of vertices that are simultaneously extremities for two partial injections.

Proposition 3. *For the uniform distribution over size- n pairs of partial injections, the probability*

$$\mathbb{P}\left(|\mathbf{extr}(f) \cap \mathbf{extr}(f')| \geq \frac{\sqrt{n}}{4(r-1)}\right)$$

is super-polynomially small (of the form $\mathcal{O}(e^{-c\sqrt{n}})$ for some $c > 0$).

Proof. Let f and f' be partial injection on $[n]$. By Proposition 2 and Eq. (2), the probability that one of them has more than $4\sqrt{n}$ extremities is super-polynomially small — so we can restrict the analysis to the cases where both f and f' have at most $4\sqrt{n}$ extremities, up to a super-polynomially small error term.

Let $m = \lfloor 4\sqrt{n} \rfloor$. Let E_f and $E_{f'}$ be two sets obtained by adding uniformly at random elements of $[n]$ to $\mathbf{extr}(f)$ and $\mathbf{extr}(f')$ respectively, until $|E_f| = |E_{f'}| = m$. Note that by symmetry, and since f and f' are chosen independently, both

E_f and $E_{f'}$ are uniform and independent size m subsets of $[n]$. Moreover, since $\mathbf{extr}(f) \subseteq E_f$ and $\mathbf{extr}(f') \subseteq E_{f'}$, we have

$$\mathbb{P}\left(|\mathbf{extr}(f) \cap \mathbf{extr}(f')| \geq \frac{\sqrt{n}}{4(r-1)}\right) \leq \mathbb{P}\left(|E_f \cap E_{f'}| \geq \frac{\sqrt{n}}{4(r-1)}\right).$$

It suffices therefore to show that, super-polynomially generically, the intersection of two m -element subsets of $[n]$ has less than $\frac{\sqrt{n}}{4(r-1)}$ elements. Let $X(n, m, k)$ be the number of pairs of m -subsets whose intersection has size k . Then

$$X(n, m, k) = \binom{n}{k} \binom{n-k}{m-k} \binom{n-m}{m-k}.$$

Therefore the probability that the intersection has size k is

$$\mathbb{P}(|E_f \cap E_{f'}| = k) = \frac{X(n, m, k)}{\binom{n}{m}^2} = k! \binom{m}{k}^2 \cdot \frac{(n-m)!^2}{n!(n-2m+k)!}.$$

Note that $\frac{(n-m)!^2}{n!(n-2m+k)!} < (n-m)^{-k}$, that $\binom{m}{k} < 2^m$. Let $\alpha = \frac{1}{4(r-1)}$. Then

$$\mathbb{P}(|E_f \cap E_{f'}| \geq \alpha\sqrt{n}) = \sum_{k=\alpha\sqrt{n}}^m \mathbb{P}(|E_f \cap E_{f'}| = k) < 2^{2m} \sum_{k=\alpha\sqrt{n}}^m \frac{k!}{(n-m)^k}.$$

Moreover $k \mapsto \frac{k!}{(n-m)^k}$ is decreasing for $k \leq m$ (for n large enough), so we have

$$\mathbb{P}(|E_f \cap E_{f'}| \geq \alpha\sqrt{n}) < 2^{2m} m \frac{(\alpha\sqrt{n})!}{(n-m)^{\alpha\sqrt{n}}} < 2^{8\sqrt{n}} 4\sqrt{n} \left(\frac{\alpha\sqrt{n}}{n-4\sqrt{n}}\right)^{\alpha\sqrt{n}}.$$

This concludes the proof since the dominant term is of the form $n^{-\frac{\alpha}{2}\sqrt{n}}$.

3.2 From partial injections to Stallings graph

Notice that if (Y, v) is a Whitehead descriptor, the definitions of the functions $\mathbf{negative}(-, Y, v)$ and $\mathbf{positive}(-, Y, v)$ make sense for all r -tuple of size n partial injections, even if they do not form a (cyclically reduced) Stallings graph. We will use the following combinatorial bounds to establish Theorem 1.

Lemma 2. *Let (Y, v) be a Whitehead descriptor and let $\mathbf{f} = (f_a)_{a \in A} \in \mathcal{I}_n^r$. If $v \in \bar{A}$, we let $f_v = f_v^{-1}$. Then we have*

$$\begin{aligned} |\mathbf{negative}(\mathbf{f}, Y, v)| &\leq \sum_{a \neq v} |\mathbf{extr}(f_v) \cap \mathbf{extr}(f_a)|, \\ |\mathbf{positive}(\mathbf{f}, Y, v)| &\geq \mathbf{sequence}(f_v) - \sum_{a \neq v} |\mathbf{extr}(f_v) \cap \mathbf{extr}(f_a)|. \end{aligned}$$

Proof. Recall that a vertex p in $\mathbf{negative}(\mathbf{f}, Y, v)$ has an incoming v -edge and all its incoming edges have labels in Y . Since $\bar{v} \notin Y$, it follows that $p \in \mathbf{extr}(f_v)$. Moreover, if $a \notin Y$ and $a \neq \bar{v}$ (there exists such an a since $|Y| \leq 2r - 2$), p has no incoming a -edge, so $p \in \mathbf{extr}(f_a)$. This establishes the first inequality.

Similarly, if $v \in A$ and p is the initial vertex of a sequence of f_v (and hence a v -extremity), and if in addition p is not an a -extremity for any $a \neq v, \bar{v}$, then $p \in \mathbf{positive}(\mathbf{f}, Y, v)$. Therefore, if $\mathbf{begin}(f_v)$ denotes the set of initial vertices of sequences of f_v , we have

$$\mathbf{begin}(f_v) \setminus \bigcup_{a \neq v, \bar{v}} \mathbf{extr}(f_v) \cap \mathbf{extr}(f_a) \subseteq \mathbf{positive}(\mathbf{f}, Y, v),$$

and the announced inequality follows since $|\mathbf{begin}(f_v)| = \mathbf{sequence}(f_v)$.

If $\bar{v} \in A$ we consider instead the set of final vertices of sequences in $f_{\bar{v}}$. \square

Proof of Theorem 1. Let \mathcal{D}_n be the set of r -tuples of size n partial injections which define a cyclically reduced Stallings graph, and let \mathcal{E}_n be the set of r -tuples \mathbf{f} of size n partial injections which fail to satisfy $|\mathbf{positive}(\mathbf{f}, Y, v)| > |\mathbf{negative}(\mathbf{f}, Y, v)|$ for some Whitehead descriptor (Y, v) . By Proposition 1, we want to show that $\mathcal{E}_n \cap \mathcal{D}_n$ is super-polynomially negligible within \mathcal{D}_n .

Since \mathcal{D}_n is generic in the full set of r -tuples of partial injections, namely \mathcal{I}_n^r (see Section 2.3), Lemma 1 shows that we only need to show that \mathcal{E}_n is super-polynomially generic in \mathcal{I}_n^r .

For each Whitehead descriptor (Y, v) , let $\mathcal{E}_n(Y, v)$ denote the set of r -tuples $\mathbf{f} \in \mathcal{I}_n^r$ such that $|\mathbf{positive}(\mathbf{f}, Y, v)| \leq |\mathbf{negative}(\mathbf{f}, Y, v)|$. Then \mathcal{E}_n is the (finite) union of the $\mathcal{E}_n(Y, v)$ and it suffices to prove that each $\mathcal{E}_n(Y, v)$ is super-polynomially generic in \mathcal{I}_n^r .

For a fixed Whitehead descriptor (Y, v) , Lemma 2 shows that

$$\mathbb{P}(\mathcal{E}_n(Y, v)) \leq \mathbb{P}\left(\mathbf{sequence}(f_v) \leq 2 \sum_{a \neq v} |\mathbf{extr}(f_v) \cap \mathbf{extr}(f_a)|\right).$$

We observe that if $|\mathbf{extr}(f_v) \cap \mathbf{extr}(f_a)| < \frac{1}{4(r-1)}\sqrt{n}$ for each $a \in A$, $a \neq v, \bar{v}$ and $\mathbf{sequence}(f_v) > \frac{1}{2}\sqrt{n}$, then $2 \sum_{a \neq v} |\mathbf{extr}(f_v) \cap \mathbf{extr}(f_a)| < \frac{1}{2}\sqrt{n} < \mathbf{sequence}(f_v)$, so that $\mathbf{f} \notin \mathcal{E}_n(Y, v)$. Therefore, by considering the complements of these properties, we see that $\mathbb{P}(\mathcal{E}_n(Y, v))$ is at most equal to

$$\mathbb{P}\left(\mathbf{sequence}(f_v) < \frac{1}{2}\sqrt{n}\right) + \sum_{a \neq v} \mathbb{P}\left(|\mathbf{extr}(f_v) \cap \mathbf{extr}(f_a)| > \frac{1}{4(r-1)}\sqrt{n}\right).$$

This concludes the proof since each of the summands is super-polynomially small by Propositions 2 and 3. \square

Theorem 1 is stated for the uniform distribution on *cyclically reduced* Stallings graphs. One may wonder if a similar result holds for the uniform distribution on Stallings graph. We show the following.

Corollary 1. *Strict Whitehead minimality is polynomially, but not super-polynomially, generic for the uniform distribution over Stallings graphs.*

Proof sketch. As per the proof of Theorem 1, a random reduced graph super-polynomially generically satisfies $|\mathbf{positive}(\mathbf{f}, Y, v)| > |\mathbf{negative}(\mathbf{f}, Y, v)|$ for any Whitehead descriptor (Y, v) . However, for H to be strict Whitehead minimal, $\Gamma(H)$ must also be cyclically reduced, that is, vertex 1 must be of valency at least 2. If vertex 1 has valency 1, it must be an extremity for one letter and isolated for all other letters. The probability it happens is $\Theta(n^{-(r-1)-\frac{1}{2}})$, by Proposition 2 and Eq. (1), concluding the proof. \square

4 The word-based distribution

Let $k \geq 2$ be a fixed integer. We discuss the genericity of strict Whitehead minimality for the subgroups generated by a random k -tuple of cyclically reduced words and we show the following.

Theorem 2. *For the uniform distribution over k -tuples of cyclically reduced words of length at most n , strict Whitehead minimality is exponentially generic.*

4.1 Shape of the Stallings graph

The following elementary statement combines results established in [1, 8] and in [2, Sec. 3.1].

Proposition 4. *Let $\alpha \in (0, 1)$ and $0 < \beta < \frac{1}{2}\alpha$, let $\mathbf{h} = (h_1, \dots, h_k)$ be a tuple of elements of \mathcal{R}_n and let H be the subgroup generated by \mathbf{h} . Then, exponentially generically,*

- $\min |h_i| > \lceil \alpha n \rceil$ and the prefixes of the h_i and h_i^{-1} of length $\lfloor \beta n \rfloor$ are pairwise distinct
- the Stallings graph $\Gamma(H)$ consists of a central tree of height $\lfloor \beta n \rfloor$ – whose vertices can be identified with the prefixes and suffixes of length at most $\lfloor \beta n \rfloor$ of the h_i – and of k outer loops, one for each h_i , of length $|h_i| - 2\lfloor \beta n \rfloor$, connecting the leaves of the central tree.

Proposition 4 describes the typical shape of a Stallings graph under the word-based distribution: as β can be taken arbitrarily small and α arbitrarily close to 1, an overwhelming proportion of the vertices are in the outer loops, and in particular have valency exactly two.

4.2 Counting the occurrences of short factors

If u is a word over an alphabet B , we denote by $Z_n(u)$ the function that counts the occurrences of u as a factor in a word in B^n .

Lemma 3. *Let B be a finite alphabet with $k \geq 2$ letters and let $u \in B^m$. Then for any $\varepsilon > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P} \left(\left| Z_n(u) - \frac{n}{k^m} \right| \geq \varepsilon n \right) \leq e^{-cn}.$$

	a	\bar{a}	b	\bar{b}
φ_a	1	-	3	2
$\varphi_{\bar{a}}$	-	3	1	2
φ_b	1	3	2	-
$\varphi_{\bar{b}}$	1	3	-	2

w	b	a	\bar{b}	\bar{a}	b	b	b	a	a	\bar{b}	a	b	a	\bar{b}	a
$\varphi(w)$	\bar{b}	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{2}$	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{1}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{1}$

Fig. 2. An example of the encoding used in the proof of Lemma 4. The word w above is encoded using the construction associated with the pattern $u = a\bar{b}$: a is always encoded by 1, \bar{b} by a 2 and the inverse of the first letter, \bar{a} , by a 3. An occurrence of u always corresponds to an occurrence of $\mathbf{12}$ in $\varphi(w)$, but the opposite is not true: there are false positives, which are always preceded by a 3. Note also that an occurrence of $\mathbf{312}$ does not always correspond to a false positive.

Proof. For $i \in [n - m]$, the probability $X_n^{(i)}$ that u is a factor at position i in a random word of length n is k^{-m} . For each $0 \leq \ell < m$, let $Z_n^{(\ell)}(u) = \sum_j X_n^{(mj+\ell)}$, for $0 \leq j \leq \lfloor \frac{n-\ell}{m} \rfloor$. Each $Z_n^{(\ell)}(u)$ is the sum of independent random variables since there is no overlap in the portions of the length n word considered. Therefore $Z_n^{(\ell)}(u)$ follows a binomial law of parameters k^{-m} and $\lfloor \frac{n-\ell}{m} \rfloor$: by Hoeffding's inequality [7], it is centered around its mean value which is equivalent to $\frac{n}{mk^m}$, and it satisfies $\mathbb{P}\left(\left|Z_n^{(\ell)}(u) - \frac{n}{mk^m}\right| > \frac{\varepsilon}{m}n\right) \leq e^{-c_\ell n}$ for some $c_\ell > 0$ and for each n large enough. The announced result follows from the fact that $Z_n(u) = Z_n^{(0)}(u) + \dots + Z_n^{(m-1)}(u)$. \square

Lemma 4. *Let $u = u_1u_2$ be a reduced word of length 2. There exists a constant $c > 0$ such that, for n large enough,*

$$\mathbb{P}\left(Z_n(u) > \frac{n}{(2r-1)^2} + \varepsilon n\right) \leq e^{-cn} \text{ and } \mathbb{P}\left(Z_n(u) < \frac{n(2r-2)}{(2r-1)^3} - \varepsilon n\right) \leq e^{-cn}$$

Proof. We first consider the case where $u_1 \neq u_2$. The idea is to use Lemma 3 via an encoding of reduced words. For every $a \in \tilde{A}$, let φ_a be a bijective map from $\tilde{A} \setminus \{\bar{a}\}$ to $[2r-1]$. Let φ be the map from the set of reduced words to $\tilde{A} \times [2r-1]^*$ defined for every reduced word $z = z_1 \cdots z_n$ by

$$\varphi(z) = (z_1, \varphi_{z_1}(z_2))\varphi_{z_2}(z_3) \cdots \varphi_{z_{n-1}}(z_n).$$

Observe that for every $n > 0$, φ is a bijection from \mathcal{R}_n to $\tilde{A} \times [2r-1]^{n-1}$, which is computed by an automaton with outputs: the states are the elements of \tilde{A} and for every $a \in \tilde{A}$ and $b \neq \bar{a}$, there is a transition from a to b on input b with output $\varphi_a(b)$. Moreover, the uniform distribution on \mathcal{R}_n is obtained by choosing z_1 uniformly in \tilde{A} , z uniformly in $[2r-1]^{n-1}$, and taking $\varphi^{-1}(z_1, z)$.

We now choose particular functions φ_a : for every $a \neq \bar{u}_1$, we choose $\varphi_a(u_1) = 1$. This way every occurrence of u_1 (except possibly for the first letter of w), is encoded by a 1 (note that the 1s provided by $\varphi_{\bar{u}_1}$ do not encode an occurrence of u_1). We also require that $\varphi_{u_1}(u_2) = 2$ and $\varphi_a(\bar{u}_1) = 3$ for every $a \neq u_1$: thus

every occurrence of $u = u_1u_2$ in w translates to an occurrence of 12 in $\varphi(w)$, and every occurrence of \bar{u}_1 translates to a 3 in $\varphi(w)$. See Figure 2 for an example.

Then for any t , we have $\mathbb{P}(Z_n(u) > t + 1) \leq \mathbb{P}(Z_{n-1}(12) > t)$ (the value $t + 1$ in the left-hand side of the inequality corresponds to the possibility of an occurrence of u at the leftmost position). Since the pattern 12 is taken in $[2r - 1]^*$ equipped with the uniform distribution, we apply Lemma 3 to get the first inequality.

Observe that counting occurrences of 12 over-estimates the number of occurrences of u . More specifically, a false positive occurs if, and only if, the said occurrence of 12 is preceded by a 3 in $\varphi(w)$. Hence, the number of false positives is bounded above by the number of occurrences of 312 in $\varphi(u)$. Therefore $\mathbb{P}(Z_n(w) < t) \leq \mathbb{P}(Z_{n-1}(12) - Z_{n-1}(312) < t)$. We can use Lemma 3 to get the second inequality (since $\frac{1}{(2r-1)^2} - \frac{1}{(2r-1)^3} = \frac{2r-2}{(2r-1)^3}$).

The case $u = u_1u_1$ is handled in the same fashion, except that we have to set $\varphi_{u_1}(u_1) = 2$ instead of 1. \square

Remark. The statement of Lemma 4, and even a slightly stronger statement, can also be obtained using the theory of Markov chains: a reduced word can be seen as a path in a specific Markov chain – where the set of states is \hat{A} , and there is a transition from a to b with probability $\frac{1}{2r-1}$ whenever $a \neq \bar{b}$. The result in Lemma 4 then follows from [11, Thm 1.1]. We chose instead to give the elementary and self-contained presentation above.

4.3 Proof of Theorem 2

Let $\alpha \in (0, 1)$, $\beta \in (0, \frac{\alpha}{2})$ and $\varepsilon > 0$ be real numbers, to be chosen later. Let $Z_{n,\alpha,\beta}$ be the set of k -tuples $\mathbf{h} = (h_1, \dots, h_k)$ of reduced words of length at most n , such that $\min |h_i| > \lceil \alpha n \rceil$ and the prefixes of the h_i and h_i^{-1} of length $\lfloor \beta n \rfloor$ are pairwise distinct.

For each word h of length greater than $2\lfloor \beta n \rfloor$, let $\text{mid}(h)$ be the factor of h obtained by deleting the length $\lfloor \beta n \rfloor$ prefix and suffix.

Now let (Y, v) be a Whitehead descriptor and let H be the subgroup generated by $\mathbf{h} \in Z_{n,\alpha,\beta}$. We denote by Y^c the complement of Y . The central tree of $\Gamma(H)$ has at most $2k\beta n$ vertices, and the outer loops of $\Gamma(H)$ are labeled by the $\text{mid}(h_i)$. The vertices in these loops all have valency 2. Any one of these vertices is in $\text{negative}(\Gamma(H), Y, v)$ if and only if it has an incoming v -edge and an outgoing y -edge for some $y \in Y^c \setminus \{v\}$. Let $N = (Y\bar{v} \cup v\bar{Y}) \setminus \{v\bar{v}\}$. Then the number of negative vertices in the outer loops is equal to the number of occurrences of elements of N as factors in the $\text{mid}(h_i)$. That is:

$$\text{negative}(\Gamma(H), Y, v) \leq \sum_{i=1}^k \sum_{xy \in N} Z_{|\text{mid}(h_i)|}(xy) + 2k\beta n.$$

By Proposition 4, $Z_{n,\alpha,\beta}$ is exponentially generic. Moreover, the map $h \mapsto \text{mid}(h)$ turns the uniform distribution on words in \mathcal{R}_ℓ ($\ell > \alpha n$) into the uniform distribution on $\mathcal{R}_{\ell-2\lfloor \beta n \rfloor}$: indeed, if $u \in \mathcal{R}_{\ell-2\lfloor \beta n \rfloor}$, then $\mathbb{P}(\text{mid}(h) = u) = (2r -$

$1)^{2\lfloor\beta n\rfloor}$, which does not depend on u . Therefore, exponentially generically, we have

$$\mathbf{negative}(\Gamma(H), Y, v) \leq 2k(|Y| - 1) \left(\frac{1}{(2r - 1)^2} + \varepsilon \right) (1 - 2\beta + \varepsilon)n + 2k\beta n.$$

Similarly, a loop vertex is in $\mathbf{positive}(\Gamma(H), Y, v)$ if it has an incoming x -edge with $x \in Y \setminus \{v\}$ and an outgoing y -edge with $\bar{y} \in Y^c$: if $P = (Y \setminus \{v\})\bar{Y}^c \cup Y^c(\bar{Y} \setminus \{\bar{v}\})$, then the number of positive vertices in the outer loops is equal to the number of occurrences of elements of P as factors in the $\mathbf{mid}(h_i)$. That is, exponentially generically,

$$\begin{aligned} \mathbf{positive}(\Gamma(H), Y, v) &\geq \sum_{i=1}^k \sum_{xy \in P} Z_{|\mathbf{mid}(h_i)|}(xy) \\ &\geq 2k(|Y| - 1)(2r - |Y|) \left(\frac{2r - 2}{(2r - 1)^3} - \varepsilon \right) (\alpha - 2\beta)n. \end{aligned}$$

In order to conclude, we only need to show that we can choose α , β and ε such that $(2r - |Y|) \left(\frac{2r - 2}{(2r - 1)^3} - \varepsilon \right) (\alpha - 2\beta) > \left(\frac{1}{(2r - 1)^2} + \varepsilon \right) (1 - 2\beta + \varepsilon) + \frac{\beta}{|Y| - 1}$. This is possible by continuity, since the limits of these two quantities when $(\alpha, \beta, \varepsilon)$ tends to $(1, 0, 0)$ are respectively $(2r - |Y|) \frac{2r - 2}{(2r - 1)^3} \geq \frac{4}{3} \frac{1}{(2r - 1)^2}$ and $\frac{1}{(2r - 1)^2}$.

This establishes that if H is generated by a k -tuple of reduced words, then exponentially generically $\mathbf{positive}(\Gamma(H), Y, v) > \mathbf{negative}(\Gamma(H), Y, v)$ for each Whitehead descriptor. The same exponential genericity holds for k -tuples of cyclically reduced words in view of Lemma 1 and the discussion at the end of Section 2.3. Together with Proposition 1, this concludes the proof since a subgroup generated by a tuple of cyclically reduced words has a cyclically reduced Stallings graph.

Remark. If one start with a k -tuple of reduced words instead of cyclically reduced words, there is a non-negligible probability that there exists a letter $a \in \hat{A}$ such that every word starts with a and ends with \bar{a} , in which case the graph is not cyclically reduced. Asymptotically, this happens with probability that tends to $(\frac{1}{2r})^{2k-1}$.

5 Application to random generation

Our main theorems can be used to sample Whitehead minimal subgroups for the two distributions, in **linear average time**. The idea is to use a rejection algorithm by repeatedly drawing a Stalling graph until the result is Whitehead minimal.

The average complexity of such an algorithm is $\frac{c_n}{p_n}$, where p_n is the probability of success, and c_n is the average complexity of an attempt. By Theorem 1 and Theorem 2, $p_n \rightarrow 1$ in both settings.

For both distribution, a graph can be generated in linear average time: using the algorithm described in [3] for the graph-based distribution; relying on

Touikan’s algorithm [17], which is linear on average as remarked in [4, Theorem 4.1], for the word-based distribution.

Observe that strict Whitehead minimality can be tested in linear time using Proposition 1 and the fact that there are finitely many Whitehead descriptors. If this test fails the subgroup can still be Whitehead minimal, so we use the polynomial algorithm of [15] to test Whitehead minimality. The key point is that, as a consequence of Theorem 1 and Theorem 2, this algorithm is called with super-polynomially small probability, and therefore the average complexity of the Whitehead minimality test is linear.

References

1. G. N. Arzhantseva, A. Yu. Ol’shanskiĭ. Generality of the class of groups in which subgroups with a lesser number of generators are free. *Mat. Zametki*, 59:489-496, 638, 1996.
2. F. Bassino, A. Martino, C. Nicaud, E. Ventura, P. Weil. Statistical properties of subgroups of free groups. *Random Struct. Algorithms*, 42:349-373, 2013.
3. F. Bassino, C. Nicaud, P. Weil. Random generation of finitely generated subgroups of a free group. *Internat. J. Algebra Comput.*, 18:375-405, 2008.
4. F. Bassino, C. Nicaud, P. Weil. Generic properties of random subgroups of a free group for general distributions. In *23rd Intern. Meeting on the Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AQ, pp. 155-166, 2012.
5. Ph. Flajolet, R. Sedgewick. *Analytic combinatorics*. Cambridge Univ. Press, 2009.
6. S. M. Gersten. On Whitehead’s algorithm. *Bull. Amer. Math. Soc.*, 10:281-284, 1984.
7. W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13-30, 1963.
8. T. Jitsukawa. Malnormal subgroups of free groups. In *Computational and statistical group theory*, Contemp. Math., vol. 298, pp. 83-95. Amer. Math. Soc., 2002.
9. I. Kapovich, A. Miasnikov, P. Schupp, V. Shpilrain. Generic-case complexity, decision problems in group theory, and random walks. *J. Algebra*, 264:665-694, 2003.
10. I. Kapovich, P. Schupp, V. Shpilrain. Generic properties of Whitehead’s algorithm and isomorphism rigidity of random one-relator groups. *Pacific J. Math.*, 223:113-140, 2006.
11. P. Lezaud. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, 8:849-867, 1998.
12. R. C. Lyndon, Paul E. Schupp. *Combinatorial group theory*. Springer, 1977. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 89.
13. A. D. Miasnikov, A. G. Myasnikov. Whitehead method and genetic algorithms. In *Computational and experimental group theory*, Contemp. Math., vol. 349, pp. 89-114. Amer. Math. Soc., 2004.
14. Y. Ollivier. *A January 2005 invitation to random groups*, *Ensaos Matemáticos*, vol. 10. Soc. Bras. de Matemática, 2005.
15. A. Roig, E. Ventura, P. Weil. On the complexity of the Whitehead minimization problem. *Internat. J. Algebra Comput.*, 17:1611-1634, 2007.
16. J. R. Stallings. Topology of finite graphs. *Invent. Math.*, 71:551-565, 1983.
17. N. W. M. Touikan. A fast algorithm for Stallings’ folding process. *Internat. J. Algebra Comput.*, 16:1031-1045, 2006.